

SUMMARY MINUTES

MEETING OF THE ORTHOPAEDIC AND REHABILITATION DEVICES ADVISORY PANEL

OPEN SESSION

June 2 and 3, 2004

**Gaithersburg Marriott Washingtonian
Gaithersburg, MD**

Meeting of the Orthopaedic and Rehabilitation Devices Advisory Panel

Attendees

June 2, 2004

Chairperson

Michael J. Yaszemski, M.D., Ph.D.

Executive Secretary

Janet L. Scudiero, M.S.

Members

Maureen A. Finnegan, M.D.

John S. Kirkpatrick, M.D.

Sanjiv H. Naidu, M.D., Ph.D.

Deputized Voting Members

Marcus P. Besser, Ph.D.

Brent A. Blumenstein, Ph.D.

Fernando G. Diaz, M.D., Ph.D.

Choll W. Kim, M.D., Ph.D.

Jay D. Mabrey, M.D.

Consumer Representative

Kleia Luckner, J.D., M.S.N.

Industry Representative

Sally L. Maher, Esq.

FDA Representative

Celia M. Witten, M.D., Ph.D.

CALL TO ORDER

Executive Secretary Janet L. Scudiero, M.S. called the meeting to order at 9:57 a.m. She stated that Marcus P. Besser, Ph.D., Brent A. Blumenstein, Ph.D., Fernando G. Diaz, M.D., Ph.D., Choll W. Kim, M.D., Ph.D., and Jay D. Mabrey, M.D. were appointed to temporary voting status for this meeting. She then read the conflict of interest statement. Full waivers were granted to John S. Kirkpatrick, M.D. and Jay D. Mabrey, M.D. for their interests in firms that could be affected by the panel's recommendations. The agency took into consideration other matters concerning Drs. Finnegan, Kim, Kirkpatrick, and Mabrey, all of whom reported current or past interests in firms at issue but in matters not related to the day's agenda. They could participate fully in the panel's deliberations. Ms. Scudiero noted that the next panel meeting dates are tentatively scheduled for August 12–13 and December 2–3, 2004.

Panel Chair Michael Yaszemski, M.D., Ph.D., stated that the purpose of meeting was to make recommendations concerning a PMA for the DePuy Charité Artificial Lumbar Disc for spinal arthroplasty in skeletally mature patients with degenerative disc disease (DDD) at one level from L4 to S1. He asked the panel members to introduce themselves, after which he noted that the members present constituted a quorum.

Barbara Zimmerman, Chief, Orthopedic Devices Branch, updated the panel on several submissions since the previous panel meeting in December 2003. The Wyeth INFUSE bone graft, discussed at the November 2002 meeting, was approved April 30, 2004. The bone graft is indicated for treating acute, open tibial shaft fractures that have been stabilized with IM nail fixation after appropriate wound management. The PMA is now owned by Medtronic Sofamor Danek.

Five non-panel track PMAs were approved since the December meeting. The Stryker OP-1 Putty, which was collaboratively reviewed with CDER and CBER, was approved April 7, 2004. The material is indicated for use as an alternative to autograft in compromised patients requiring revision posterolateral (intertransverse) lumbar spinal fusion, for whom autologous bone and bone marrow harvest are not feasible or are not expected to promote fusion.

The Oxford Meniscal Unicompartamental Knee Phase 3, manufactured by Biomet, was approved April 21, 2004. The device is indicated for use in patients who have osteoarthritis or avascular necrosis limited to the medial compartment of the knee and intended to be implanted with bone cement.

Two human demineralized bone matrix (DBM)-based bone void fillers were approved this spring: The Exactech Resorbable Bone Paste was approved February 27, 2004, and the Wright Allomatrix Putty was approved March 5, 2004. The Kyphon KyphX Bone Cement, polymethylmethacrylate cement for pathological fracture of the vertebral body, was approved April 1, 2004.

Reclassification of the interbody fusion devices (cage) was discussed at the December 2003 Panel meeting. The agency is currently drafting a special controls guidance document and a reclassification proposed rule for the device.

OPEN PUBLIC HEARING

Ms. Scudiero read the agency's statement on transparency of the device approval process. She noted that seven patients and family members wrote to FDA requesting approval of the Charité disc. In addition, the agency received abstracts of two presentations that took place during the Spine Week 2004 meeting this week. One abstract contained information on complications related to cage devices, and the other was a case report of problems related to polyethylene wear debris.

Steven Kurtz, Ph.D., Exponent, Philadelphia, presented a retrieval analysis (funded by Medtronic Sofamor Danek) on an explanted Charité disc. It showed that low levels of oxidation were measured at the surface, but they were not associated with a reduction in mechanical properties. Surface damage was observed, but the direction of cracks was not associated with magnitude and distribution of oxidation. Regions of high-tensile stress in the model corresponded to observations of cracking and rim damage.

John Pelloza, M.D., a spine surgeon in Texas and an investigator for the Maverick Total Disc Replacement device, stated that the Charité total disc replacement device raises clinical, materials, fixation, and kinematics issues. The U.S. trial produced superior clinical outcomes to the BAK control, but published studies show reoperation rates of 5 to 20 percent and complications in more than 10 percent of the cases. In no joint replacement has ultra high molecular weight polyethylene (UHMWPE) stood the test of time for 40 years; UHMWPE degradation can cause severe inflammatory reactions that lead to bone-metal loosening. In addition, fixation is not adequate and will predictably fail. The implant in the U.S. study has no porous coating for bony ingrowth, so the implant is susceptible to loosening, subsidence, and

migration. In addition, there are reports of dislocation. The best results are attributed to surgeon skill and expertise; this will not be sustained if the device is approved. Moreover, even if the implant is perfectly placed without fixation complications, its location places increased force on the facet joint, leading to facet degeneration. The concept of a posterior fusion to rescue a failed implant is not likely to work predictably. Most implants will need to be removed using an anterior approach, which is potentially life threatening.

David W. Polly, Jr., M.D., Professor and Chief of Spine Surgery, University of Minnesota, said that disc arthroplasty technology represents a paradigm shift in the field's expectations of spinal implant performance. Failures of the discs, however, will present significant revision challenges. Implanting surgeons must understand the potentially life threatening difficulties associated with revision procedures. The idea of joint registries has appeal: The Swedish joint registry has been helpful in early identification of problems. In the U.S., however, HIPAA constraints make prospective data collection difficult. The implantation of the device must be trained for intensively, and lessons learned must be disseminated widely.

SPONSOR PRESENTATION

Bill Christianson, Vice President, Clinical and Regulatory Affairs, DePuy Spine, Raynham, MA, introduced the sponsor presenters and noted that several consultants were in attendance. The Charité disc has been available in Europe since 1987.

Paul C. McAfee, M.D., Towson Orthopedic Associates, summarized the history of the device design. The uncoated version has a 16-year track record. He described the rationale for the design and its biomechanics, noting that the device has five sizes of footprints to match the normal spine. A review of adverse event reports reveals infrequent complications: one case of osteolysis in Australia, and one case of fractured UHMWPE in Europe. These are technical complications, not problems inherent in the device itself. The one device that fractured at 9.5 years was successfully revised.

Published reports (Lemaire 2002, Rachis 2002, and David 2000 and 2003) on more than 315 patients with at least 12 months of follow-up indicate generally good results. Surgeon experience is important in successful outcomes. The U.S. study has received the benefits of worldwide refining of the device's indications. The device design mimics the motion of an intact disc. Most long-term data is with uncoated end plates, and FDA wanted a study on the most

recent design with the most experience, so that is the device under review. Minimal adverse events have been reported. Specialized surgical training is required, and appropriate patient selection is vital.

Bryan W. Cunningham, M.Sc., Director, Spinal Research, Union Memorial Hospital, Baltimore, MD, summarized the mechanical testing and wear simulation, *in vitro* biomechanical modeling, and *in vivo* animal modeling. The submitted preclinical testing includes the preliminary testing before 1994 and the FDA's recommended supplementary testing following ASTM draft standards WK453 (static and fatigue testing) and WK454 (wear debris testing). The mechanical testing found that the Charité disc has high compressive strength properties to address physiological demands and provides sufficient resistance to permanent compressive deformation under prolonged fatigue loading conditions. The device provides sufficient fatigue strength for the intended use. It generates low levels of UHMWPE wear particles compared with other joint arthroplasty devices.

In vitro biomechanical modeling quantified the multidirectional flexibility properties of the Charité device compared with interbody cages and pedicle screws plus cages. The Charité disc outperformed the two other devices in the percentage of axial rotation, flexion/extension, and lateral bending that remained compared to an uninstrumented functional spinal unit. The device reestablishes kinematics to the operative functional spinal unit.

Cadaveric and functional animal studies found that the disc restored motion at the operative level. No evidence of an acute neural or systemic histopathological response due to wear debris was found in either the functional animal study or the rabbit neurotoxicity study, although the latter study did find granulation tissue and a chronic histiocytic reaction.

Scott Blumenthal, M.D., Texas Back Institute, lead investigator, presented the clinical results. The data show that the Charité disc is safe and effective and at least as good as anterior lumbar interbody fusion (ALIF) with the BAK cage. The clinical study was a randomized trial involving 15 U.S. centers, each of which had 5 training cases. Patients were randomized 2:1, and target enrollment was 194 Charité disc cases and 97 BAK cage cases; 75 cases were training cases. The noninferiority study compared the safety and effectiveness of the Charité disc to ALIF with BAK cage for treatment of single-level DDD.

Dr. Blumenthal listed the key inclusion and exclusion criteria and reviewed the proposed indication for use. Primary endpoints were Oswestry Disability Index (ODI) improvement of ≥ 25

percent from baseline to 24 months, no additional surgery at the treated level, no major complications (defined as major vessel injury, neurological damage, or nerve root injury), and maintenance of neurological status from baseline to 24 months. All four criteria had to be met for a patient to be considered a success. Secondary endpoints included ODI score, pain measured by the VAS, the SF-36 score, change in disc height, device displacement, range of motion, duration of hospitalization, and patient satisfaction.

Methods to minimize bias included validated patient self-report questionnaires, independent review of neurological results and radiographs, and assigning treatment the day of surgery. The sponsor chose BAK cage as the optimum control group in consultation with FDA; it was the accepted state-of-the-art technology at the time. The surgical approach was the same, and morbidity was similar. At all intervals, follow-up was >90 percent. The patient demographics did not differ significantly except that the BAK group had higher BMI and lower preoperative activity level than patients receiving the Charité disc.

In an analysis comparing all randomized patients (205 Charité disc and 99 BAK cage), the Charité disc and BAK cage had similar adverse event profiles. Few device-related adverse events occurred in either group. Five Charité disc patients experienced device displacement or migration; four remained stable and in place, and one required additional fixation. Charité disc patients had lower revision rates than BAK cage patients. The sponsor concluded that the Charité disc is safe compared with BAK cages.

Dr. Blumenthal then reviewed the effectiveness data, focusing on the intent to treat population (182 Charité disc patients and 85 BAK cage patients), which excludes subjects not complete through 24 months. Results for range of motion and disc height were comparable. Pain relief was better in Charité disc patients than in BAK cage patients, and the percentage of nonresponders was similar or lower than that reported in prior literature for all treatment modalities. No evidence implicates facet joints. Mean ODI, VAS, and SF-36 scores were comparable in both groups. The Charité disc outperformed BAK cage in patient satisfaction. Radiographic results showed near physiologic range of motion on flexion and extension for Charité disc and good maintenance of disc height. No device displacements occurred in the BAK cage group, but five occurred in the Charité disc group. Heterotopic ossification occurred in 6 Charité disc patients at 12 months and 11 patients at 24 months. The Charité disc is at least equivalent to the BAK cage.

Looking at the entire randomized population, 62 percent of Charité disc patients and 49 percent of BAK cage patients showed = 25 percent improvement in ODI score, which was a statistically significant difference. VAS scores for pain favored the Charité disc group at 6 weeks, 3 months, and 12 months; SF-36 scores and patient satisfaction favored Charité disc.

Training cases had longer surgery duration and a higher rate of adverse events, but they generally outperformed the study patients on measures of pain and the ODI. The sponsor concludes that the device is effective.

George DeMuth, M.S., Stat-Tech Services, LLC, presented an overview of the statistical issues. He noted that FDA had pointed out that the sponsor had not conducted a sensitivity analysis for the different patient group comparisons. The sponsor's subsequent sensitivity analysis showed results well below the required 15 percent threshold; the results strongly support the noninferiority claim.

Bill Christianson described the sponsor's proposed physician training program, which will take place in coordination with the Spine Arthroplasty Institute in Cincinnati, OH. New users will attend the training before distribution of the device. The primary training will be augmented by geographically dispersed regional training centers. The training will consist of 12 modules that include hands-on training. The course will be provided using written and CD-based versions of the module.

FDA PRESENTATION

Sergio de del Castillo, biomedical engineer, Orthopedic Devices Branch and lead reviewer for the Charité Artificial Disc PMA, described the device and presented the FDA's preclinical review. The device for which the company is seeking approval contains only uncoated endplates. Because the sponsor already summarized the mechanical testing that was conducted, he did not elaborate on the testing any further but noted that although the mechanical testing results appear to represent the expected physiological loads and range of motion, the correlation of these results to the clinical performance of the device is not known. He deferred to Dr. Graham's presentation for an account of the wear debris testing.

Mr. de del Castillo then presented the agency's clinical review. The purpose of the sponsor's study was to evaluate the safety and effectiveness of the Charité disc and compare it to the BAK cage. The study sought to demonstrate that the Charité disc performed at least as well

as the BAK cage within a noninferiority margin, or delta, of 15 percent. Although the BAK cage may be implanted using either an open anterior or posterior approach, for the purposes of the study, all control subjects were implanted with the BAK devices only with an open anterior approach. The Charité device is implanted using only this approach.

The proportion of Charité disc and BAK cage subjects experiencing at least one adverse event was essentially equal. However, some adverse events were reported for a higher percentage of Charité disc subjects than BAK cage subjects, including infection, abdominal events, device-related events, and severe or life-threatening events. In addition, 7.3 percent of Charité disc subjects experienced device-related adverse events, compared with 4.0 percent of BAK cage subjects. A greater percentage of Charité disc subjects than BAK cage subjects experienced back or lower extremity pain; neurological events, such as numbness, motor deficit, or nerve root injury; and additional surgery at the index level. The rate of adverse events was higher in the training subjects group than in the randomized subjects, a finding that may be attributed primarily to the slightly higher rates of prosthesis-related events and additional surgeries at the index level. The training subjects were not included in the assessment of safety.

The agency's assessment of the primary and secondary endpoints was based on the "Completers" group, a subset of all randomized subjects who were evaluated at the 24-month time point. The group contains 86 percent and 79 percent of all randomized Charité disc and BAK cage subjects, respectively. The success rates for the Charité disc and the BAK cage groups are 64 percent and 58 percent, respectively. Although these rates differ slightly from what the company presented, the success rate is within a noninferiority margin, or delta, of 10 percent of the BAK cage success rate. The study has therefore demonstrated the noninferiority of the Charité disc to the BAK cage. The only component where there is a statistically observed difference between the two study groups is the ODI score.

Because one of the principle theoretical advantages of disc replacement devices is the preservation of segmental motion, FDA considered the correlation between success and range of motion observed. FDA compared success and failure rates at 24 months for Charité disc subjects with their range of motion data. Subjects experiencing range of motion in the 5 to 7 degrees range were more likely to be successful than subjects experiencing different ranges of motion. However, the association of range of motion with success is not statistically significant.

In summary, the Charité disc demonstrated noninferiority to the BAK cage with respect to the primary endpoint. The numbers of adverse events in the Charité disc and BAK cage groups was equivalent, with a higher rate of incidence in only a few categories for the Charité disc group. The Charité disc was able to maintain pain and function out to 24 months. Some subjects reported only some pain relief; and a few experienced no change or an increase in pain. Finally, it is unclear how range of motion is related to the clinical outcomes, if at all.

Jianxiong Chu, Ph.D., M.A.S., statistician, Division of Biostatistics, provided the agency's statistical summary. After reviewing the study methodology and results for the primary and secondary endpoints, he noted that excluding patients biased the results. The BAK cage group had a high proportion of noncompleters who were excluded from the ITT analysis. All the sponsor's claims of the Charité disc's superiority to BAK cage with respect to the secondary endpoints were based on the unadjusted P values without a prespecified plan to control the study-wide Type I error rate. To demonstrate that the Charité disc provides a benefit at an earlier time point after implantation than BAK cage, time to sustained benefit should be compared between the two groups. The statistical analysis provides evidence that Charité disc is at least as good as BAK cage (using a noninferiority margin, or delta, of 10 percent), except under the worst-case scenario. However, the sponsor's sensitivity analyses may be biased against the control BAK cage group. No formal claim regarding secondary endpoints should be made without multiplicity adjustment to control the study-wide Type I error rate. Adverse events might be under reported in the current submission. The most recent data, including discontinued and overdue patients, need to be submitted and analyzed.

Jove Graham, Ph.D., engineer and reviewer, CDRH, presented data on testing and evaluation of wear debris. It is unknown what the biological response to wear particles of the size generated by the device will be in the human spine. No spinal disc literature or data exist for direct comparison, and limitations exist to what clinical conclusions can be drawn by comparing spinal disc testing to hip or knee replacements. The average wear rate is 0.11 mg/mc; the wear debris is mostly submicron, and the average diameter is 0.21 to 1.49 μm . The wear rate is smaller than for most reported wear rates for hip and knee replacements. Particles of UHMWPE implanted into the spinal region can cause epidural fibrosis, macrophage reaction, and transient upregulation of IL-6, and particles of the same size can elicit different reactions in different parts of the body. No reactions specific to spinal cord or cerebrospinal fluid were seen. Preclinical

testing has done a good job of characterizing wear behavior of the device. However, we cannot establish the safety and effectiveness of a spinal device by comparing preclinical results to hip or knee devices, and we cannot validate results of any wear test simulator until we have explanted retrievals for comparison. Wear-induced osteolysis is a long-term complication that may not be observed until 10 or 15 years of follow-up.

PANEL PRESENTATIONS

John S. Kirkpatrick, M.D., clinical panel reviewer, summarized the goals of and principles underlying disc replacement and examined how well the literature has dealt with those principles. He then discussed how the PMA addressed those principles. He noted that normal unconstrained motion has been studied using cadaveric models, motion profiles, and testing before and after disc placement. Anterior column support and normal biomechanics, especially as to how the facets are affected, are not thoroughly addressed in the literature. The PMA should be commended for its extensive report. The mobility data were unconvincing. Anterior support was well demonstrated in the PMA.

Wear data to 10 million cycles were presented, but the literature usually reports to 50 million cycles. It would have been good to have coupled flexion/extension with lateral bending. It is unclear whether the testing caused debris; the specimens contained grooves in the line of the motion direction. Other issues included incomplete data on osteointegration, heterotopic ossification, and facet degeneration and a lack of clarity as to the sponsor's measure of neurological status. Stratification among indications groups may have improved understanding of results. The follow-up intervals were well defined, but they may not have been of adequate length. The sponsor has failed to demonstrate absence of adjacent segment degeneration. In addition, it is not clear that 50 percent of a representative range of motion is "near physiologic."

Brent A. Blumenstein, Ph.D., panel statistical reviewer, stated that he basically agreed with the FDA statistical review. The sponsor's analysis is flawed; even so, the product appears to meet noninferiority criteria. The goal is to identify the best characterization of the noninferiority outcome. The term "intent-to-treat" is inaccurate and should be "analysis-by-arm"; likewise, the term "data set" is preferable to "population." The sponsor's definition of the ITT population is incorrect because it deletes randomized patients; the FDA statistician also made this point.

Dr. Blumenstein presented a brief overview of randomized controlled trial principles, emphasizing that deletions from arms erode stochastic equivalence. Deletions based on post randomization events are particularly onerous because they are more likely related to the intervention, such as side effects or intervention implementation issues. The primary outcome should be defined for all possible contingencies. The trial's quantitative measures are therefore problematic.

A lower significance level for the final analysis should be considered. Correction for interim analysis would decrease final criterion to just under 0.05. The analysis-by-arm meets the noninferiority criterion (Charité = 55.6 percent, BAK = 45.5 percent) using the Blackwelder test ($p < .0001$). The 90 percent CI is 0.1 to 20.2.

The bottom line is that the sponsor demonstrated success under all canonical analyses. Sensitivity testing demonstrated success under all but the worst-case scenarios. Were the trial to be designed today, it would be important to look hard at a failure time primary endpoint (e.g., failure-free survival). The trial would capture time and handle missing data better.

PANEL DISCUSSION

Panel members asked the sponsor for clarification as to the learning curve, factors underlying revisions, methodology for determining the absence of wear debris, rationale for the 6-month sacrifice point in the baboon study, type of polyethylene being used in the device, likelihood that the devices will last for 40 years in younger patients, impact of the device on adjacent segment disease, indications for anterior revision, and center of rotation of the device.

Question 1: Please comment on the results of the wear debris testing and particulate analysis.

The panel agreed that the testing was adequate, but members expressed several concerns, including the need for data on aged specimens. The sponsor may be able to provide data from total joint replacements. Another concern was that the disc may turn into a synovial-like device after the procedure, and long-term follow-up—more than 2 years—may be necessary to see the effects of wear debris. Other concerns included the possibility for cracks and fragmentation and the possible long-term response of nerve tissue to chronic inflammation.

Question 2: Please discuss the clinical significance of non-device-related pain, wound infections, device-related additional surgery at index level and any other adverse events seen in the trial.

The complication rate was surprisingly low. Although non-device-related pain complications and infections seemed higher with the Charité device than with BAK, and additional surgery was four times that occurring with BAK, the device is a moving device and more prone to failure, so the comparison may not be fair.

Question 3: Please comment on the interpretation of [the pain relief] findings.

The panel concluded that because pain is subjective and difficult to assess. In addition, the Charité group was more active and had lower BMI.

Question 4: Please comment on the sponsor's claim that the Charité permits "near physiological movement with up to 15 degrees bending in flexion/extension and a similar degree of lateral bending and axial rotation to the natural disc."

The panel believed that the flexion and extension results obtained with the Charité device were within the normal range, but the results do not include rotation. The link between range of motion and clinical improvement shows a trend, but it was not demonstrated.

Question 5: Do the clinical data in the PMA provide reasonable assurance that the device is safe?

The panel concurred that over the study period, the device is safe. Questions on long-term safety, i.e., longer than 24 months, remain.

Question 6: Do the clinical data in the PMA provide reasonable assurance that the device is effective?

The panel concurred that the device is effective.

Question 7: If you recommend approvability for this PMA, do you recommend a post-approval study? If so, please discuss what types of endpoints would be useful for an updated label and recommend the duration of such a study.

The panel agreed that post approval follow-up is necessary. Dr. Kirkpatrick provided a list of suggested endpoints, many of which were incorporated into the panel's recommended conditions of approval. Many questions can be answered with existing data. Panel members recommended examining European data and thought 5 years of follow-up was appropriate.

OPEN PUBLIC HEARING

Stephen Hochschuler, M.D., Chair, Texas Back Institute, and a board member of the Spine Arthroplasty Society, read a statement of the society's position on educational and training goals for spine surgeons interested in new arthroplasty technologies. Comprehensive formal training should be followed by proctorship at the training surgeon's hospital for his or her first

case(s). This approach offers significant long-term advantages for patients, surgeons, industry, and hospitals. The society intends to develop guidelines for educational programs and identify training centers with adequate facilities and staffing who can provide proctorship. Certification only verifies that the surgeon has completed training,

Dr. A. van Ooij, Maastricht, Netherlands, presented data on complications of the Charité disc prosthesis in 49 patients. Numerous early and late complications occurred in the group, including anterior migration, facet joint degeneration, subsidence, subluxation, breakage of the metal wire around the core, and degenerative scoliosis. Most patients who had additional operations had posterior fusions without removing the prosthesis. The malfunctioning prosthesis continues to cause pain. Good placement and sizing seems difficult; even patients with good surgery had problems. The center of rotation of the Charité disc is too anterior and is not like a normal disc. Wear will be a problem in the future. Revision is dangerous and sometimes impossible, and the claim of preventing adjacent disc degeneration is not substantiated.

Pamela Adams, Orthopedic Surgical Manufacturer's Association (OSMA), reviewed the regulatory definitions of safety, effectiveness, and valid scientific evidence. She emphasized that the standard is reasonable assurance, balancing benefits with risks.

FDA SUMMATION

Dr. Witten clarified the difference between a condition of approval and a not-approvable recommendation. She also clarified that it is the agency's option as to whether to take the PMA back to panel in the future.

SPONSOR SUMMATION

Sponsor representatives noted that the company had conducted a 24-month randomized controlled trial in accordance with FDA's guidance document. Long-term follow-up from Lemaire and David provides valid scientific evidence. Post approval studies are an appropriate and accepted means to develop long-term safety data, and the company is amenable to a 5-year follow-up study. However, wear testing to 50 million cycles is excessive. The testing already submitted represents 80 years of significant bends while lifting 20 kg. The device has been on the market since 1987 and is fully bio-mechanically characterized. A robust clinical study

demonstrated safety and effectiveness. The sponsor requested that the panel recommend approval.

VOTE

Ms. Scudiero read the three voting options. The first panel motion was for not approvable; this motion was voted down, six against disapproval and two for disapproval. The second motion was for approval with conditions. The panel voted unanimously to recommend approval of the device with the following conditions:

1. The sponsor should continue to follow all patients enrolled in the IDE study until the last enrolled subject reaches the 2-year time point.
2. All patients who are treated with the Charité disc should be provided with documentation of the name of the device, the device's serial or lot numbers, an identification number, the name of the surgeon, where the surgery took place, the date of surgery, and a telephone number for reporting adverse events.
3. The sponsor should collect additional data on wear debris. Wear debris testing should use combinations of flexion/extension and lateral bending motions (without axial rotation) to 10 million cycles.
4. FDA should require that the company make training available with the understanding that certification will be left to state licensing boards and credentialing committees.
5. FDA and the sponsor should discuss the following conditions of approval to come to a mutually agreeable course of action. This discussion will consider whether each item should be addressed pre- or post market:
 - a. Provide mobility testing data or complete references.
 - b. Provide a rationale for "normal biomechanics," including demonstration or data that facet joint strains/stresses are comparable to those of the control group patients.
 - c. Provide an adequate rationale for not testing the biological response to submicron UHMWPE wear particles.
 - d. Clarify the neurological grading scale and how statistics were applied to that measure.
 - e. Stratify results by indication group, especially for the two groups with facet joint changes in preoperative studies, into both fusion and disc replacement groups.
 - f. Define patients with range of motion of 0 to 5 degrees due to loss of disc function as failures
 - g. Consider whether it is appropriate to use axial imaging to compare preoperative facet degeneration and degeneration at 24 months.

- h. Provide radiographic evaluation of adjacent segment degeneration at the preoperative and 24-month time points, as well as through the time period described in Condition #1.

POLL

When asked to explain the rationale for their votes, panel members generally indicated that they believed the sponsor provided sufficient data to assure the members of the device's safety and efficacy. The panel's concerns were allayed by the conditions of approval.

ADJOURNMENT

Drs. Yaszemski and Witten thanked the participants, the panel, and the sponsor. Dr. Yaszemski adjourned the meeting at 5:22 p.m.

Meeting of the Orthopaedic and Rehabilitation Devices Advisory Panel

Attendees

June 3, 2004

Chairperson

Michael J. Yaszemski, M.D., Ph.D.

Executive Secretary

Janet L. Scudiero, M.S.

Members

Maureen A. Finnegan, M.D.

John S. Kirkpatrick, M.D.

Kinley Larntz, Ph.D.

Sanjiv H. Naidu, M.D., Ph.D.

Deputized Voting Members

Marcus P. Besser, Ph.D.

Choll W. Kim, M.D., Ph.D.

Jay D. Mabrey, M.D.

Michael B. Mayor, M.D.

Consumer Representative

LeeLee Doyle, Ph.D.

Industry Representative

Sally L. Maher, Esq.

FDA Representative

Celia M. Witten, M.D., Ph.D.

CALL TO ORDER

Executive Secretary Janet L. Scudiero, M.S., called the meeting to order at 8:01 a.m. and announced upcoming tentatively scheduled meetings. She stated that Marcus P. Besser, Ph.D., Choll W. Kim, M.D., Ph.D., Michael B. Mayor, M.D., and Jay D. Mabrey, M.D., were appointed to temporary voting status for the duration of the meeting. She then read the conflict of interest statement. Full waivers were to Drs. Kim and Mabrey for their interests in firms that could be affected by the panel's recommendations. The agency took into consideration certain matters regarding Drs. Finnegan, Kim, Kirkpatrick, and Mabrey, all of whom reported current or past interests in firms at issue but in matters not related to the day's agenda. They could participate fully in the panel's deliberations.

Panel Chair Yaszemski noted for the record that the panel members present constituted a quorum and stated that the purpose of the meeting is to provide recommendations to the FDA on an OSMA-initiated reclassification proposal to reclassify mobile bearing knee (MBK) joint prostheses and on a draft hip guidance document submission (GDS) on performance criteria for hip joint prostheses. This is the first industry group-prepared GDS.

OPEN PUBLIC HEARING

Ms. Scudiero read the FDA's statement on transparency of the device approval process. She also noted for the record that the American Academy of Orthopaedic Surgeons had sent a statement to the agency expressing support for reclassifying MBKs into class II.

Stephen J. Peoples, V.M.D., M.S., Worldwide Vice President, Clinical and Regulatory Affairs, DePuy, stated that the sponsor claims that strong evidence exists for the safety and effectiveness of MBKs. However, the petition fails to justify a general reclassification. Most literature presented involves a single MBK design, the LCS. The Accord MBK design had almost a 50 percent failure rate in clinical use.

The petition presents limited data on a limited number of designs. Large variation exists in the revision rate from design to design. IDE data are included in the review, but most studies involve short follow-up in small populations. No data are provided on 63 percent of the MBK designs identified in the petition. The sponsor's proposed special controls do not address issues involving polyethylene wear and kinematics, including femorotibial stability and "spin out."

The petition under consideration reviews a single total MBK and a single MBK unicompartmental device that the petitioners claim represents safety and effectiveness of MBK designs in general. Minute design differences affect device function. The petition does not meet the requirements for reclassification.

John Fisher, Director, Institute of Medical and Biological Engineering and Pro-Vice-Chancellor, University of Leeds, United Kingdom, noted that not all MBKs are the same. MBKs are complex systems. Motion at individual wear interfaces is design dependent and cannot be predicted from whole joint kinematics. Evidence increasingly indicates that lower contact stresses and larger wear areas increase surface wear and wear debris; however, that is not the case in the LCS model, which has unidirectional motion and much lower wear than fixed bearing knees (FBKs). MBKs are prone to scratching and third-body damage, particularly on the tibial interface. Clinical and laboratory studies show that debris from fixed-bearing knees is larger and less reactive than debris from hips. It is speculated that MBK debris is more like hip debris. This may be the case for multidirectional designs due to cross-shear, as found in hip and fibril/particle fragmentation; research, however, has shown that unidirectional motion produces larger less reactive debris.

With regard to whether wear questions can be addressed through preclinical simulator tests, two standards for knee simulators exist: force control and displacement control. Results have been mixed in comparison of MBKs and FBKs. Results are design and test-method dependent. McEwen developed a special combined force and displacement controlled testing mode for rotating platform (RP) mobile bearings, but it is not an ISO standard test and is not available in other simulator systems. Wear volume cannot be easily determined from clinical measurements. Not all MBKs are alike, and the impact of their many design variables is not understood. The reclassification petition does not address the effects of the design variables on the performance of MBKs. The special controls proposed the petition (currently used for class II FBKs), therefore, cannot ensure that the various designs of MBKs are safe and effective.

Douglas Dennis, Adjunct Professor, Department of Biomedical Engineering, University of Tennessee, stated that the reclassification petition assumes that all available MBK designs will demonstrate similar efficacy and safety. MBKs, however, differ in underside motion patterns, and they have a potential for premature wear and periprosthetic osteolysis from increased underside wear. The wear debris created with an FBK is larger and less reactive than

that created with MBKs, creating the potential for more osteolysis. In addition, multidirectional motion accelerates UHMWPE wear. Minor differences in kinematics can produce major differences in wear magnitudes. Long-term clinical results of multidirectional wear designs are not yet available. The FDA should proceed with caution in grouping all MBKs as class II devices.

Cris Ruddlesdin, FRCS, Consultant Orthopaedic Surgeon, Barnsley District General Hospital, UK, described his experience with the Rotaglide rotating/gliding MBK. The device has been on the market since 1988 and more than 20,000 have been implanted worldwide. He has implanted the MBK in 119 patients since 2000; average length of follow-up is 2 years. In those patients, one gross dislocation occurred. Data from other patient cohorts show various complication rates. To avoid certain complications, correct ligament balancing and correct flexion/extension gaps are important, whether the device is fixed or mobile. There is no difference in the level of difficulty to implant a fixed or MBK. Insert dislocations are not a clinical issue, and gross dislocations are not happening at a significant rate. Whether the knee is fixed or mobile bearing, the final arbiter to a good fit is the surgeon's tactile feel during trial reduction.

Tom Ferring, M.D., an orthopedic surgeon from Charlotte, NC, and consultant for DePuy, expressed concern about the potential for wear debris with new MBK designs. It is counter intuitive to conclude that a variety of new MBK designs are equivalent to proven designs. In orthopedics, we rarely see failures before five years. New MBK implants must be reviewed with standard scientific methods.

Barry Soros, M.D., an orthopedic surgeon, Little Rock, AR, stated that he was an original clinical investigator for the LCS knee. The patients were served well by that IDE. In his experience, more than half of complications patients experience are due to surgeon technical errors. MBKs have a steep learning curve, so the IDE/PMA process is good for patients.

David Fitzpatrick, biochemical engineer, University College, Dublin, focused on the issue of special controls. It is clear that outcomes are dependent on management of soft tissues and operative technique. Postoperative stability is critical. The clinical history of MBKs shows that revision is a common outcome. Preclinical tools do not have the ability to predict clinical or kinematic performance.

INDUSTRY PRESENTATION—OSMA

Toni R. Kingsley, Ph.D., Warsaw, IN, explained OSMA's purpose and stated that the association is requesting reclassification of MBK devices from class III into class II. The petition requests reclassification of total and unicompartmental MBKs both cemented and uncemented. MBKs have been on the market for 25 years and now represent the third generation of devices. Forty-six designs are available worldwide; six designs are approved in the U.S. FDA asked that the reclassification petition address all MBKs, rather than just subcategories. The petitioner contended that MBKs and FBKs do not differ in clinical performance, and that the clinical performance of the various MBK designs and surgical techniques are the same.

James B. Stiehl, M.D., Clinical Associate Professor of Orthopaedic Surgery, Medical College of Wisconsin and Columbia and St. Mary's Hospital, Milwaukee, WI, summarized published and unpublished clinical data on MBKs. The unpublished data are from OSMA companies supporting the petition and come from IDE trials or international clinical outcomes studies. The published data are from a comprehensive literature review from 1997 to July 2, 2002, and selected review articles. The literature suggests that MBK devices perform similarly to well-designed FBK devices in survivorship and clinical function. Current IDE and international outcomes studies suggest that other MBK designs are clinically successful and comparable to fixed bearing designs. Osteolysis and patellar complications are minimal. The potential benefit of this technology is improved long-term clinical performance and longevity.

Peter S. Walker, Ph.D., New York Medical Center, New York, NY, a knee designer for Stryker and Zimmer, reviewed the risks and proposed special controls. Of tens of thousands of MBK knees, there are only about 385 MDR reports. The most common adverse events are pain or swelling, bearing fractures, loosening, and metal/poly separation. Other known risks include dislocation and subluxation, and wear may be a concern.

All design features were evaluated for the MBKs referenced in the petition, and each knee was assigned to one or more MBK type (multidirectional platform, meniscal bearing, etc.). The potential biomechanical advantages and disadvantages were determined for each mobile bearing type, and they were used to establish the associated risks. Finally, special controls that addressed each risk were determined and listed for each mobile bearing characteristic (e.g., platform, meniscal bearing, rotational stops, and congruency).

The risks and special controls were divided into two groups: (1) risks and special controls that are common to FBKs and MBKs and for which there are no special issues related to MBK design features and (2) risks and special controls that have specific considerations when applied to MBKs as compared with the same special control applied to FBKs. Current ASTM and ISO standards could be modified for MBKs.

In summary, MBK risks are well understood and are similar to FBKs risks. Special controls (guidance documents, ASTM and ISO standards, regulations, etc.) for these risks either exist and are commonly used in industry or can be adapted for any unique characteristic of a specific mobile bearing design. A new FDA special controls guidance document that describes each test and test parameters is needed. OSMA believes that special controls, when combined with the general controls, will be sufficient to provide reasonable assurance of the safety and effectiveness of MBKs.

Greg Maislin, M.S., M.A., Biomedical Statistical Consulting, Wynnewood, PA, described the methodology of and results from a metaanalysis of data. Randomized clinical trials comparing MBKs to FBKs are largely not available in the literature, so methods of meta-analysis appropriate for observational studies were used. Two meta-analyses were performed: (1) clinical outcomes and (2) implant survival.

Good evidence indicates that survival is similar for MBKs and FBKs. The variability in implant survival was similar to the variability in revision rates in sets of FBKs that were studied. The most important consequence of wear is increased revision rates, and the revision rates of MBKs can be predicted from fixed bearing counterparts.

The meta-analysis found that the MBKs and FBKs are similar in both effectiveness and survival. MBK characteristics (e.g., cemented vs. non-cemented) did not demonstrate significant differences in clinical outcomes or implant survival.

FDA PRESENTATION

Peter Allen, reviewer, Orthopedics Devices Branch, presented an overview of the device reclassification process. Special controls guidance documents are created by FDA to describe acceptable methods for controlling the risks identified for a given device type. They convey FDA's current thinking about a specific device type and provide recommendations on how to address the issues presented in the special controls guidance document. A company need

only demonstrate that its class II device meets the recommendations of the special controls guidance document to receive FDA clearance for marketing. MBKs are postamendment class III devices, and they require an approved PMA prior to marketing. FDA has approved three MBK PMAs. One is for a total MBK device, and two are for unicompartmental MBK devices.

This petition is split into two groups of MBK designs. The first consists of a “total” knee design, which contains patella, femoral, and tibial components and is intended to replace the entire knee joint. The second consists of a unicompartmental design; it contains only femoral and tibial components and is intended for replacement of either the medial or lateral compartment of the knee. Both device types are available in a multitude of design variations. Many variables can affect the design of a total MBK. The reclassification of the currently approved devices would potentially provide for the reclassification of these various design variables, many of which are incorporated into the currently approved devices. Although much fewer in number, various combinations of design variables can also go into the development of unicompartmental knee devices. Mr. Allen reviewed the proposed indications for use for the total and unicompartmental MBK devices.

OSMA included a bibliography of more than 230 published references in support of the preclinical and clinical issues in this petition. The preclinical issues addressed include evaluation of device kinematics, wear of the mobile bearings, and device biomechanics. The sponsor summarized clinical data on a series of 48 studies; data presented for each study included study design, demographics, safety, effectiveness, and survivorship. Most studies focused on devices already approved for use in the U.S. (i.e., the LCS and Oxford MBK devices). The data underscore the strong influence of the technical performance of the operation on the long-term success of a knee device. Properly aligned knee replacements that have restored ligament balance appear to have survival rates of 10 years or greater, irrespective of bearing mobility. The data indicate that when provided with medial-lateral stabilization, MBKs provide equivalent results to FBKs.

The sponsor’s information on adverse events included data gathered from searches of FDA’s MDR program, reports from the published literature, and data from various manufacturers from their FDA approved clinical trials. Most of this information relates to the DePuy LCS devices and the Biomet Oxford unicompartmental devices. The MDR reports in particular relate specifically to the DePuy LCS devices. The three most common adverse events

cited in the MDR database for the LCS knee were pain (with swelling), fractured bearings, and loosening, respectively. The patient-related adverse events are fairly typical of the type of events one might see with any total joint replacement procedure, and the device-related adverse events are consistent with the types of complications often seen with FBKs. However, there appears to be a tendency to see a greater number of bearing dislocations, subluxations, and impingement with MBKs. OSMA has proposed using ASTM and ISO standards, as well as standards in existing guidance documents, to control for device-related risks including bearing wear, bearing fracture, bearing dislocation, bearing subluxation, impingement, instability, and component loosening. However, these standards apply to FBKs and not to MBKs.

It is well known that successful implantation of MBKs is highly technique sensitive. Without proper attention to soft-tissue balancing, instability of the implanted joint is a real risk. To minimize this risk, the sponsor suggests that special attention be given to providing appropriate instructions for use of the device in the product labeling. The sponsor believes surgeon training and detailed surgical techniques that include instructions for proper soft tissue balancing will provide reasonable assurance of safety and effectiveness. No specifics were given for the recommended training, but it appears that these would be of the same type currently provided for FBKs. This approach, along with wear testing, was recommended to control against risk of prosthesis or soft-tissue impingement.

The only risk identified as unique to unicompartmental knees was that unicompartmental devices require intact anterior and posterior cruciate ligaments. To mitigate the risk of these devices being implanted in patients without functional cruciate ligaments, the sponsor has recommended product labeling and surgeon training in the proper surgical technique.

PANEL REVIEWS

Michael B. Mayor, M.D., panel clinical reviewer, said that he believes the state of the art is comparable for fixed and mobile bearing total knees. They provide some of the most predictable and cost-effective interventions available. Many of the considerations regarding wear are common to fixed and mobile bearings. Unintended motion and wear have emerged as significant factors. It is not clear that MBK designs are a source of excessive risk with regard to wear. Stability is another concern.

Do these devices expose the public to unnecessary risk? The risks are being addressed. With the means available to FDA, including performance and test standards, literature, and FDA evaluation of devices for approval, it seems prudent to recommend reclassification of MBKs. In addition, development of special controls and appropriate guidance is needed.

Kinley Larntz, Ph.D., panel statistical reviewer, observed that metanalysis is hard work. Similarity in the context of much variation is easy to achieve. It does not mean “no difference”—it means we do not have enough evidence or that the literature is quite scattered and published for various reasons. The observed differences actually are an understatement of the variation that exists.

The metanalyses were done in a fixed-effects context. The tables show clear, statistically different variations in, for example, the percentage of outcomes rated as “excellent.” However, random components need to be accounted for in any measure of variation that is given. A true random effects analysis would do that. The survival analysis itself appears to take no account of individual follow-up time in studies. It should be related to time, and it is unclear why it was not. Although it was not demonstrated that there is no difference in survival, it is likely that it is essentially the same for the two types of devices. Also, no metanalysis was done on adverse event rates; this is significant because some of the studies had much higher adverse event rates. In addition, only three PMAs have been approved for the devices; that is not a big experience set. Also, OSMA should have identified the studies in the metanalysis using a numbering system or some other means. In sum, it is hard to draw conclusions from the data presented. A lack of statistical difference does not mean that no difference exists.

Dr. Witten clarified the goal of reclassification. She stated that the designs that have been presented are potentially eligible for class II. The agency would like to hear a discussion about whether enough is understood about the ability of the proposed special controls, such as preclinical tests, to predict performance (relative safety and effectiveness) so that risks can be controlled.

PANEL DISCUSSION

The panel discussed whether all MBKs can be treated alike, or whether unicompartmental and tricompartmental should be considered separately. Panel members concurred that they should be considered separately. They noted that with the Oxford knee,

preclinical testing would not have caught the problems that arose during clinical use. Experience derived during the clinical trial has made that device safer to use. The panel also discussed whether it was appropriate to treat all MBKs as a group because there are subtle differences between designs. No one set of special controls will necessarily cover all MBKs. It was noted that reclassification would apply to the devices that are currently on market and that the agency is good at looking at applications and determining whether a device is substantially equivalent.

PANEL QUESTIONS

Question 1: Do you believe the proposed classification definitions for the . . . device configurations recommended for reclassification adequately describe the devices? If not, what changes in the definitions do you recommend?

The panel believed that the proposed definitions are broad and that they adequately define these devices. Concerns include the fact that “the patellar device” needs more clarification as to whether it is mobile or not and as to joint loading. The definition for unicompartmental is broad, but it makes testing more difficult.

Question 2: Do you believe the risks to health of the following device configurations proposed for reclassification are adequately described? If not, what additional risks do you believe should be included?

The panel concurred that the completeness of risks to health with respect to unicompartmental MBKs are adequately described. Multicompartmental knees need additional special controls. Unicompartmental MBKs may need to be separated from total MBKs.

Question 3a. Dislocation and subluxation of MBK components have been cited as common complications in the literature.

- i. Do you believe appropriate special controls have been identified to adequately address these risks?
- ii. If not, what additional controls, if any, do you recommend to address these risks?

The panel stated that although these complications are most common, they are not common in themselves. They are primarily a result of technique errors. Controls are adequate to identify mechanical problems in the device itself to address dislocation risk. Training is a necessity.

3b. A reduction in wear is often cited as a theoretical advantage of MBKs over FBKs. However, this has not been consistently demonstrated clinically, and it is not clear how well preclinical wear testing of MBKs correlates to the clinical situation. In addition, the potential for third body wear appears greater (due to the fact you have 2 moving interfaces instead of one). Currently, the state of development of knee simulator wear testing has not yet been standardized or clinically validated for all design types of MBKs, and therefore may not be applicable for all of the various MBK types identified in this petition.

- i. In light of the fact that wear appears to be, in part, design dependent, do you believe appropriate controls have been identified to adequately address the risk of wear (i.e., osteolysis, loosening) for the various MBK designs under consideration in this petition?**
- ii. If not, what additional controls, if any, do you recommend to address this risk?**

The panel agreed that the ability to characterize wear debris has improved and such controls should be available to the sponsor. Testing should look at uni- and multidirectional wear patterns. New tests may need to be developed for multidirectional wear. In the absence of a joint simulator test, postmarket wear analysis and retrieval analysis are needed. No special control is adequate to test all design configurations, but ISO and ASTM standards may establish a baseline for these devices.

3c. Labeling has been cited as a method with which to control some of the identified risks to health. The proposed labeling requirements are consistent with those generally found in current FBK package labeling. Such labeling typically includes adequate instructions for use, device description, indications for use, contraindications, adverse events, precautions, warnings, a listing of compatible components, and sterility information.

- i. What additional labeling, if any, do you recommend for these MBK devices?**

The panel noted that the effectiveness of the devices is surgical technique dependent. The implanting surgeon needs to be familiar with total knee replacement. The labeling and recommendations are appropriate. The devices should be restricted to use by people who have been adequately trained.

3d. Do you believe appropriate special controls have been identified to adequately address the risks to health for each of the above device configurations (and all 'subconfigurations')? If not, what other special controls do you recommend to address the risks presented by these devices?

The panel concurred that appropriate special controls have been identified. The mechanical and preclinical testing is good, and clinical data should be included. Wear testing should combine multiple motion modes. Uni- and multidirectional devices should be considered separately.

4. Do you believe the data presented in this petition supports the reclassification of:

- a. All total MBK prostheses identified in this petition? If not, which types of total MBKs do you believe are inappropriate for reclassification, and why (e.g., they have insufficient information and/or special controls)?**
- b. All unicompartmental MBK prostheses identified in this petition? If not, which types of unicompartmental MBK's do you believe are inappropriate for reclassification, and why?**

The panel was not in agreement. Most believed that data supported reclassification of all MBK devices presented in the petition, but others expressed concern about reclassifying unicompartmental devices, and at least two panel members opposed reclassification of both types of devices.

CLASSIFICATION QUESTIONNAIRE AND VOTE

The panel first attempted to fill out a single classification questionnaire for both total and unicompartmental MBKs. Because the panel could not reach a consensus, they then agreed to fill out this form separately for both generic types of MBKs. The panel voted six to two to recommend that the agency reclassify total MBKs into class II. The panel recommended the following special controls: a special controls guidance document, testing guidelines, potential use of clinical data, device-specific training and labeling (to be negotiated with sponsors), and patient documentation that lists the name of the device, the device's serial or lot numbers, an identification number, the surgeon's name, the name of the hospital, the date of surgery, and a telephone number for reporting adverse events.

The panel voted five to three to recommend that FDA reclassify unicompartmental MBKs into class II. The panel recommended the same special controls as for the total MBK, emphasizing clinical data and long-term follow-up. Postmarket surveillance should track adverse events such as osteolysis, revisions, dislodgment or motion of implant, and polyethylene failure.

POLL

Panel members voting in support of reclassification believed that the sponsor had proposed adequate controls to ensure safety and efficacy of the device and to ensure proper development of the device components. Panel members voting against reclassification were concerned about inadequate clinical data and lack of comparability of MBKs. Many panel members supporting reclassification believed that clinical data are appropriate and necessary.

HIP GUIDANCE DOCUMENT SUBMISSION

OPEN PUBLIC HEARING

No comments were made.

INDUSTRY PRESENTATION

Joel Batts, OSMA, presented an overview of the Hip Guidance Document Submission (Hip GDS). The problem with hip replacement system (HRS) control groups is that the variation of the devices makes comparison difficult, creating a burden for researchers, as well as scientific limitations. The purpose of the Hip GDS is to move toward benchmark development. The Device

Forum initiated the Hip GDS with input from clinicians, FDA, and industry. It covers a range of study purposes and creates a three-point composite benchmark based on literature and clinician and scientist consensus.

A short-term benefit of the Hip GDS is that it provides clinicians, industry, and FDA with a less burdensome, more reliable method of conducting clinical trials. It also provides patients with a clearer understanding of the risks and benefits of study participation and improves confidence in conclusions from data analysis. Long-term benefits include “apples-to-apples” comparisons of study results and a foundation for updating clinical and scientific consensus as the body of knowledge grows.

Bernard Stulberg, M.D., Center for Joint Reconstruction, Cleveland Orthopedic and Spine Hospital, OH, summarized how the Hip GDS was developed. A two-step approach was used to create a valid document: review of literature and clinician and scientist consensus. OSMA reviewed 277 articles for type and frequency of complications at two years. The 1,489 complications identified were divided into four main categories: device only, operative technique only, operative technique and device, and systemic/unrelated. A consensus was developed through participation of fourteen members of the Hip Society of the American Academy of Orthopedic Surgeons.

In the Hip GDS, a study subject is considered to be successful if, at endpoint, he or she has had no device-related complications, has a Harris Hip Score (HHS) of ≥ 80 , and has not had revision surgery. Device-related complications are those in the complications list developed from the literature review. The standardized HRS clinical trial objective defines a successful study as one in which at least 95 percent of the HRS device group subjects are successful at endpoint according to the composite definition.

Joshua J. Jacobs, M.D., orthopedic surgeon, Chicago, described the Hip GDS’s implications for the clinician with regard to scientific, study logistics, and recruitment issues. The previous approach required physicians to use two or more devices, involved subjective determination of difference in treatment effects, and was based on patients’ limited access to or desire for information on the device and surgical technique. Traditional study designs require data sets from comparable patients and devices and data sets from comparable intraoperative and postoperative treatment protocols. However, bias is not eliminated even in randomized designs. The clinician knows which device is used in the patient at the time of surgery and follow-up. The

delta is arbitrary, and treatment effect differences are subjectively chosen. The timeline to detect clinically significant differences is not in accord with regulatory timelines, and it is difficult to establish homogeneity between groups.

HRS clinical studies need to be integrated within the clinician's practice. This involves many considerations, such as data collection, IRB review, and HIPAA requirements. A significant number of patients are required, and the likelihood of attrition increases as the number of subjects needed increases. Recruitment is increasingly difficult: Patients are more proactive and self-informed, and they increasingly request specific devices and operative techniques. It is harder to create a control group because clinicians have specific preferences.

The Hip GDS takes seriously the limitations mentioned above by enlisting clinician consensus based on extensive literature and clinical experience. It allows for a more standardized method of study design, review, protocol writing, data collection, and submission. Finally, it creates a reference point from which future benchmarks may be set as the body of knowledge grows.

FDA PRESENTATION

Barbara Buch, M.D., medical officer, Orthopaedic Devices Branch, FDA, explained what a GDS is and stated that FDA would not repeat the information OSMA had already presented. She said that FDA would consider the concepts presented in the Hip GDS, and also that the agency also has several concerns that it wanted panel comment on. These concerns would be the focus of her remarks.

The proposed study duration is one year, but peer-reviewed journals, the FDA, and FDA advisory panels generally require two years of follow up. The panel has often expressed that two years of follow-up is inadequate. The Hip GDS does not justify a 1-year follow-up time frame. In addition, it uses objective performance criteria (OPC) as a control. Although historical controls often are the least burdensome approach, are a standard approach, involve valid scientific evidence, and have the potential to facilitate review of the data, they also have some drawbacks. They involve one-armed observational studies, result in compromised comparative statistical inference, assume that knowledge gathered can answer new clinical questions, and assume that a review of the literature allows complete and adequately detailed records for comparison. In addition, such an approach results in temporal bias, and historical criteria applied to a new device

may not discern whether the device is inferior to current treatment. Randomized controlled trials, however, potentially compensate for unknown biases and confounding factors of a population sample. Trials need tools to mitigate bias.

Dr. Buch reviewed the primary and secondary endpoints and noted that more appropriate surrogate outcomes may be available. She noted that it was unclear which types of radiographic evidence—radiolucency, subsidence, migration, etc.—are associated with implant or patient failure. Traditionally, studies have not used subjective patient evaluations of pain and may not have a place in evaluating hip replacement systems. The FDA's biggest concern is whether the Hip GDS captures all adverse events and revisions.

Finally, Dr. Buch presented information on implant survival and revision rates from three large data sources: The Swedish Total Hip Replacement Register, the 1994 NIH Consensus, and the Dartmouth Atlas of Musculoskeletal Health Care. These indicate that revision rates continue to be low in all studies at 2, 5, and 10 years in comparison to that suggested in the Hip HGS which would allow for a potential 5% failure (revision) rate at one year. .

Phyllis Silverman, M.S., Division of Biostatistics, presented information on OPC use. The Hip GDS proposes a one-arm study using a target value that is a fixed historical control. Each patient is labeled a success or failure according to clinically defined criteria. Then the proportion of study successes is statistically compared to the target value. Delta is the margin of noninferiority, or clinically insignificant amount.

When designing a study, such as that proposed by this Hip GDS, one picks a target value and delta, sets the Type 1 error and the power, and then computes the sample size. One can also fix the sample size and delta and see what observed study success one must meet and what the power is to do so. No matter what delta is chosen, target minus delta equals minimum guarantee. Sample size increases as target value decreases. Sample size increases as delta decreases, and sample size increases as power increases. To increase power, one must decrease variability. Several examples using different deltas and study outcome success were compared to show how this would affect the success of the study.

PANEL REVIEWERS

Jay D. Mabrey, M.D., reviewed the agency's definition of least burdensome. Total hip arthroplasty is one of the most successful operations ever created, and 30-year follow-up shows

excellent results. The state of the art has changed substantially since the 1994 NIH consensus statement was written. Various total hip designs, fixation methods, and surgical techniques need to be compared with each other. Rehabilitation interventions and patient-level predictors are important. Long-term follow-up is essential to determining outcomes and pathological processes. Failures related to osteolysis and debris are identified only through long-term follow-up.

Multiple combinations of components are available today. Metal-on-metal hips raise the problem of cobalt and chromium ion concentration in blood, the significance of which is unclear. Total hip arthroplasty has evolved into a family of procedures involving many approaches; entire catalogs are devoted to new instrumentation used in these approaches.

Follow-up duration of 24 months is appropriate. Six-week and 6-month follow-up is useful for determining early complications; later follow-up can detect failures of materials and device incorporation but this requires more than 24 months. Early failure may not be evident at 12 months, particularly in older patients.

Indications are extending to younger and older patients, and patient selection is associated with race and level of income. In initial studies of a device, sponsors should consider stratification of patients. Data are more powerful with grouping, especially if there are no concurrent randomized controls. Numbers of patients may vary, depending on variables being studied.

HHSs have been validated against other measures. In addition, sponsors should consider using quality-of-life surveys, such as the SF-36, along with disease-specific surveys. The WOMAC osteoarthritic index is a possibility, too. Outcomes are affected by factors beyond the implant itself. A HHS of >90 in every case would be ideal, but 80 or better is acceptable and is a conservative approach.

Concerning postmarket studies, continued follow-up is the norm for most total joint surgeons. Routine radiographs and examinations should occur at 1- or 2-year intervals, but it can be difficult to get patients back into the doctor's office. In most cases, continued reporting is not burdensome. The U.S. total joint registry is still under development. Surgeons continue to collect data in order to have publishable research.

Hip systems present special concerns because they are modular devices that have interchangeable bearing surfaces and geometry. Devices do not always come from same

manufacturer, and surgeons can mix and match fixation and materials. It is important to strictly follow the protocol if the devices are part of an ongoing study.

Kinley Larntz, Ph.D., described his review of guidance documents that specify OPC and discussed some of the differences among them. For hip replacement, the field might be better off using historical data, because there is so much of it. Control groups can go wrong. If one has data and can do the matching, then OPC are acceptable. Multiple outcomes should be assessed. For example, what if HHSs were acceptable, but 5 percent of the patients had revisions at one year. Historical data would say that is not a good outcome. Standards for revisions, hip scores, and complications should be set separately.

PANEL QUESTIONS

Question 1: Please discuss each of the proposed benchmark criteria in the submission and, if not adequate, discuss what options would be reasonable in terms of endpoints, sample size, success rate, or any other parameters.

The panel was not in agreement as to whether a composite endpoint was most appropriate or whether several scores should be used, one of which could be a composite. It agreed that more discussion was needed between OSMA, clinicians, and FDA. It was noted that the clinical community is more comfortable with 24-month rather than a 12-month follow-up. The panel generally believed that it was acceptable to have a study that can demonstrate that the true success rate for the device is no less than 91 percent.

Question 2. Please comment on the duration of patient follow-up in the context of the proposed composite OPC for patient and study success presented in this document. Include a discussion of the time patients should be followed after treatment.

The panel did not reach a consensus. Many panel members believed that historical controls may be acceptable, and they were divided as to whether 1- or 2-year follow-up was most appropriate. Panel members noted the difficulty with patient retention through 2-year follow-up. Two-year follow-up might be most appropriate if using a composite score for success.

Question 3: Please discuss any inclusion and exclusion criteria that would be important to incorporate in the guidance.

The panel agreed that studies should use standard inclusion and exclusion criteria, such as patients who are not pregnant; have no psychiatric problems; and have no known factors affecting outcomes of total joint replacement, including BMI, activity levels, and diagnoses. Standard international exclusions should be used. The study population should mimic the general

population that would receive the implant. Sponsors should record demographic characteristics such as age, sex, race, and weight.

Question 4: Please propose and discuss any new ideas for appropriate alternative outcome measures or other surrogate endpoints to predict success in patients who may be younger, healthier, heavier, and more active than those in the historical literature reviewed.

The panel concurred that some sort of radiological follow-up and other patient satisfaction measure, such as the SF-36, WOMAC, HHS, or return to activity, was appropriate.

Question 5: Please comment on the types of questions a postmarket study may appropriately address; the duration of follow-up that would be necessary; and the amount and type of data that should be collected to answer the posed questions after device clearance or approval.

The panel generally agreed that long-term postmarket follow-up may be appropriate; it may be the only way to answer some questions. X-ray follow-up and follow-up on some adverse events may be necessary even if a 1-year endpoint is chosen.

Question 6: Based on your experience and the experience in published literature, please comment on the types of hip systems that would be amenable to the use of OPC and which are not.

The panel agreed that the OPC represent minimum requirements; less used systems, custom devices, tumor systems, and revision devices need further FDA review. One-year follow-up is most appropriate for primary joint replacement and gives the most consistent data. It is hard to know which types of future hip joint replacement devices the Hip GDS may be appropriate for.

ADJOURNMENT

Drs. Yaszemski and Witten thanked the participants, the panel, and the sponsor. Dr. Yaszemski adjourned the meeting at 3:56 p.m.

I certify that I attended this meeting of the Orthopaedic and Rehabilitation Devices Advisory Panel on June 2 and 3, 2004, and that these minutes accurately reflect what transpired.

Janet L. Scudiero, M.S.
Executive Secretary

I approve the minutes of the June 2 and 3, 2004, meeting as recorded in this summary.

Michael J. Yaszemski, M.D.
Chairperson